

Article

Semantic Network Development in L2 Spanish and Its Impact on Processing Skills: A Multisession Eye-Tracking Study

M. Gabriela Puscama 

Department of Spanish and Portuguese Studies, University of Florida, Gainesville, FL 32611, USA;
mpuscama@ufl.edu

Abstract: The goal of this project was to explore how different types of vocabulary exposure shape the connections formed in the L2 lexicon and how these, in turn, affect L2 language processing. During L2 acquisition, words are often presented in thematic lists (e.g., *food*), favoring a lexicon organized by shared features (*burger-hot dog*). However, thematic lists offer only a partial picture of how words interconnect. For example, *beer* and *football* do not share any features and do not belong strictly to the same theme (*food* and *sports*, respectively); still, they co-occur frequently and are associated in the lexicon. A multisession training study and visual world eye-tracking tests were conducted to assess how different types of vocabulary exposure impact L2 processing. Intermediate L2 Spanish learners were trained under one of two conditions, thematic lists (TL, as in textbooks) or words presented in visual scenes (VS) with vocabulary related by co-occurrence. The VS group showed significant changes in their gaze patterns, resembling the naturalistic exposure baseline group (native speakers), more than the TL group. The results are interpreted in light of the anticipatory processing literature and the strength of representations as a result of naturalistic vs. formal exposure to L2 vocabulary.

Keywords: L2 vocabulary; L2 Spanish; eye tracking; second language acquisition; co-occurrence; vocabulary training; semantic network



Citation: Puscama, M. Gabriela. 2024. Semantic Network Development in L2 Spanish and Its Impact on Processing Skills: A Multisession Eye-Tracking Study. *Languages* 9: 43. <https://doi.org/10.3390/languages9020043>

Academic Editor: Batia Laufer

Received: 16 November 2023

Revised: 13 January 2024

Accepted: 19 January 2024

Published: 26 January 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vocabulary size is one of the strongest indicators of proficiency in a second language (L2) (Meara 1996). In the L2 learning context, vocabulary size has been found to correlate with reading (Qian 2002; Hwang et al. 2020), listening (Stæhr 2009), and speaking skills (Koizumi and In'nami 2013). L2 learners are exposed to large amounts of vocabulary in the classroom; yet, even advanced learners may struggle to retrieve words during spontaneous communication. Evidence from network science suggests that this is not strictly related to the number of words but rather to the organization or *topology* of the lexicon. Lexicon topology underlies how the lexicon is searched, how words are associated, and how easily they can be retrieved, a crucial aspect of language processing (De Deyne and Storms 2014).

Many prominent L1 lexical semantics models are based on the co-occurrence distribution of words (Jackson and Bolger 2014). Co-occurrence distribution refers to the frequency with which two words appear together. From network science, we know that the way words are encountered affects how they are stored (Bybee 1998; Hills et al. 2009). If two words tend to co-occur in the input, they become strongly associated in the mind, facilitating each other's activation (Collins and Loftus 1975). The learning of co-occurrence distributions happens naturally during L1 acquisition, as words are contextualized in the input (but see Hart and Risley 1992; Huttenlocher et al. 2010 for variable L1 exposure). L2 learners may be able to exploit this co-occurrence-based L1 network, but L2 words are often encountered in environments that do not offer direct access to co-occurrence distributions. Sometimes, exposure to L2 vocabulary occurs with isolated L1 translations (e.g., *snow-nieve*) or in thematic lists (e.g., *snow-rain*). The themes approach is not necessarily wrong, because words from the same theme tend to co-occur. However, thematic lists offer only a partial

picture of the word association. For example, *beer* and *football* do not belong to the same theme (*food* and *sports*, respectively), but they co-occur frequently and are associated. Even though the thematic list practice is widely accepted in L2 teaching contexts, researchers like Finkbeiner and Nicol (2003) warned that it is not based on direct relevant evidence. The idea that thematic lists are useful for L2 learning probably comes from a series of studies on monolingual speakers, showing that L1 thematic word lists were easier to memorize than unrelated word lists (Bousfield 1953; Cofer 1966; Cohen 1963). These findings have been erroneously extrapolated to the L2 context, without considering that their L1 is already in place and that they are learning new labels, not recalling known word lists. If learners are only exposed to thematic lists, then they must work harder to make richer associations to larger themes (e.g., *leisure*, *wellbeing*) or to connect vocabulary from different themes that co-occur in reality. Under these circumstances, L2 learners may have weak L2 semantic representations (Zhao and Li 2010), as isolated vocabulary or thematic lists could hinder the formation of a rich semantic structure in the lexicon.

In fact, research into L2 vocabulary from applied linguistics and psycholinguistics perspectives points out that learning vocabulary in contextualized situations (e.g., going to the vet) enhances learning compared to thematic lists (e.g., animals) (Finkbeiner and Nicol 2003; Bolger and Zapata 2011). Bolger and Zapata (2011) exposed L2 English learners to novel vocabulary items embedded in stories, which introduced either trained vocabulary within a theme (e.g., body parts), or unrelated vocabulary within a larger situation (e.g., a visit to the doctor). The participants' learning was assessed in a word-verification task using eye tracking. Their main finding was that learners trained in complex associations (not just themes) learned vocabulary better and were faster in rejecting incorrect options in the word-verification task. Therefore, if learners are exposed to a wider range of associations in the classroom, they would be expected to develop a stronger representation of lexical items.

Outside of the L2 classroom, some learners are exposed to words in immersive natural settings. Research suggests that learning in natural settings yields gains in vocabulary (Milton and Meara 1995; Jiménez Jiménez 2010; Pizziconi 2017) and fluency (Llanes and Muñoz 2009). For example, Jiménez Jiménez (2010) compared L2 Spanish learners who were enrolled in different levels at a university and students who studied abroad in Spanish-speaking countries. He assessed L2 vocabulary breadth using an L2–L1 translation task and vocabulary depth with a word association task. Jiménez Jiménez (2010) observed that L2 learners in study-abroad contexts experienced greater gains both in breadth and depth of vocabulary compared to the classroom learners. Even though the consensus is that studying abroad is beneficial, factors such as time spent abroad, initial proficiency level, and the tests used to measure vocabulary may modulate the results (for an overview, see Zaytseva et al. 2018). The gains found in studying abroad may be due to contextualized vocabulary exposure, where co-occurrence plays a role, favoring connections in the lexicon that are efficient for communication. However, the cognitive mechanisms behind these gains are poorly understood. Even though this paper does not focus on immersion, it takes direction from the research demonstrating that the type of exposure matters for the organization of the lexicon for improving L2 processing. We start from the idea that co-occurrence is key for learners to develop rich semantic networks (Cancho and Solé 2001; Jackson and Bolger 2014). Hence, the main goal of this paper is to discern the factors that influence lexicon topology, particularly whether exposure to co-occurrence distributions is beneficial to developing an efficient L2 lexicon. But what is an efficient lexicon and how can it be tested? An efficient lexicon allows for quick word retrieval during spontaneous communication. One way to tap into lexical retrieval is by looking at anticipatory speech processing. If a lexicon network has strong representations, then learners can activate those elements in the network faster when the circumstances are conducive to anticipation.

1.1. L2 Anticipatory Processing

Some studies have found no evidence of anticipation in L2 learners (Lew-Williams and Fernald 2010; Grüter et al. 2012), while others have shown that learners are able to

anticipate upcoming information using morphosyntactic (Dussias et al. 2013; Hopp 2013) and semantic (Dijkgraaf et al. 2017, 2019) cues. Recent evidence also suggests that the generation of expectations is modulated by L2 proficiency (Connell et al. 2021). However, proficiency is not the only factor accounting for contradictory findings in the literature. L2 learners are indeed capable of anticipating information, but they only do it when it is fruitful for comprehension (Kuperberg and Jaeger 2016). This means that, for example, if a cue is unreliable, if representations in the learner's mind are fuzzy, or if the cognitive load of processing the L2 input is overwhelming, then L2 learners will not find a use in engaging in anticipation, as the cost–benefit of deploying resources is too high (Kaan 2014; Kaan and Grüter 2021).

If we then assume that anticipatory processing is indicative of a strong lexicon that allows for efficient word retrieval, then a paradigm that focuses on anticipatory processing is needed to explore the impact of vocabulary instruction in the L2 vocabulary network. One of the most widely used experimental methods to study anticipation is the visual world eye-tracking paradigm.

In their seminal study, Altmann and Kamide (1999) presented participants with a visual scene (e.g., a boy surrounded by a cake, a toy train, and a ball) as they listened to sentences (e.g., 'The boy will eat the cake', or 'The boy will move the ball'). Participants launched anticipatory looks at the upcoming object when it was constrained by the verb (e.g., 'eat' can only be followed by 'cake'), but they were not able to anticipate when the context was ambiguous (e.g., 'move' can be followed by 'cake', 'toy train' or 'ball'). Since this study, the visual world eye-tracking paradigm has been an established method to research anticipatory processing, using a variety of phonological (Ito et al. 2018), morphosyntactic (Dussias et al. 2013), and semantic cues (Dijkgraaf et al. 2017, 2019).

1.2. The Present Study

This paper aims to answer this research question: does training vocabulary in thematic lists or co-occurrence associations yield gains in L2 comprehension, as indicated by anticipatory processing? The experiment presented tested whether different types of L2 vocabulary training could approximate naturalistic exposure in establishing strong semantic representations in the learners' lexicon. Intermediate Spanish learners completed a multisession training study, and semantic connections across words in their lexicon were tested using an eye-tracking visual world paradigm during the pretest, immediate post-test, and delayed post-test. The main purpose of this study was to understand some of the mechanisms underlying naturalistic exposure that render a highly connected lexicon for efficient word retrieval during speech processing. Based on previous studies (Finkbeiner and Nicol 2003; Bolger and Zapata 2011), it was hypothesized that learners trained in contextualized vocabulary based on co-occurrence would render a stronger lexicon compared to learners trained in traditional thematic lists.

2. Materials and Methods

2.1. Participants

In this study, 40 L2 Spanish learners and 16 native Spanish speakers participated. L2 learners were recruited from intermediate-level Spanish classes at a large university. These courses were at the SPAN 3, SPAN 100, and SPAN 200 levels, which are the classes following the beginner sequence at this institution and take place before the advanced courses at the 300 and 400 levels. Native speakers were recruited through word of mouth. The L2 learners completed the full experiment detailed below (4 sessions). The native Spanish speakers served as a baseline comparison, and they only completed one in-lab experimental session. From the initial pool, 1 native speaker was excluded due to poor calibration, leaving a total of 55 participants (40 L2 learners and 15 native speakers).

Participants completed a language use and history questionnaire through Qualtrics (Qualtrics, Provo, UT), adapted from López-Beltrán Forcada (2021). The L2 learners' ages ranged between 18 and 25 years old ($M = 19.40$, $SD = 1.35$), and the native speakers were

aged 19–34 ($M = 26.53$, $SD = 4.50$). In the L2 learner group, 31 participants self-identified as female, 8 as male, and 1 as nonbinary. In the native-speaker group, 10 participants self-identified as female and 5 as male. The native Spanish speakers were born in different locations, including Colombia ($N = 6$), Mexico ($N = 2$), Peru ($N = 1$), Puerto Rico ($N = 2$), Venezuela ($N = 1$), and the United States ($N = 3$). Those who were not born in mainland US, at the time of data collection, had been living in the US between 5 months and 22 years ($M = 3.67$ yrs, $SD = 6.08$ yrs).

The L2 learners targeted in this study were late Spanish learners with little to no immersion in their L2. On the language-background questionnaire, 35 participants reported being US-born and had no Spanish-speaking caregivers. One participant reported being born in Colombia, but she had been living in the US since 3 months of age. Her parents were monolingual English speakers. Four participants indicated that they had a Spanish-speaking caregiver, and 1 of them had spent more than a month in a Spanish-speaking country. A closer examination of these participants showed that two of them were late sequential bilinguals (reported Spanish onset of acquisition = 12–14 y.o.), and the other two were early sequential bilinguals (reported Spanish onset of acquisition = 3–4 y.o.). Furthermore, these 4 participants indicated that before schooling they used English at home 90–100% of the time and almost did not use Spanish growing up. Because these four participants match the profile of heritage Spanish speakers and may skew the results, the statistical analyses were repeated with these participants and without them, and the results did not differ (Supplementary Materials, Supplementary S5). Since their data did not affect the direction or significance of findings, they were ultimately included for the robustness of analyses.

The L2 learners were divided equally into two experimental groups (Visual Scene and Thematic List) by matching participants for Spanish proficiency and working-memory scores. Proficiency was tested with a category fluency task (Baus et al. 2013; Linck et al. 2009; Supplementary Materials, Supplementary S1), which is appropriate for a vocabulary study because it taps into word retrieval based on semantic associations (Luo et al. 2010), and there is some evidence that it correlates with overall measures of proficiency (Beatty-Martínez et al. 2020). Working memory was assessed with a digit span task (Supplementary Materials, Supplementary S2). Proficiency and working-memory measures were not collected for 1 participant due to technical difficulties; thus, this participant could not be matched and was randomly placed in one of the two experimental groups. This participant is not included in the descriptive statistics provided in Table 1 ($N = 39$) but was included for the remainder of the analyses.

Table 1. Summary of proficiency and working-memory scores.

	Speaking Proficiency		Working Memory
	English	Spanish	
Visual Scene Group	12.04 (1.44)	6.04 (1.31)	9.53 (1.71)
Thematic List Group	12.01 (1.40)	6.15 (1.41)	9.70 (1.72)

From the initial participant pool, some L2 learners (3 participants in Session 1, 2 participants in Session 2) were discarded from the eye-tracking analyses due to lack of fixations (the cut off was >20% of the trials), which occurs when participants look at the center of the screen and use peripheral vision for the task. Other learners did not complete all the sessions (3 participants in Session 2, 3 participants in Session 3). All participants who completed at least one in-lab session and were not discarded due to lack of fixations were included in the analysis ($N = 37$).

2.2. Materials

2.2.1. Visual World Task

For the critical trials in the visual world, 36 target Spanish words were selected from the prompts and responses of a word association task conducted with native Spanish speakers (De Deyne and Storms 2008; De Deyne et al. 2013; De Deyne et al. 2019; Dubossarsky et al. 2017; Prolific 2014; Qualtrics 2005; Supplementary Materials, Supplementary S4). The targets selected were picturable nouns because the training had a visual component. Each target appeared in a four-word display with 1 distractor and 2 fillers. The targets were presented under two conditions, semantically related (*avión-vuelo* ‘airplane-flight’) or unrelated target distractor (*avión-hueso* ‘airplane-bone’) (Figure 1). The semantically related distractors, also appeared as targets on different lists, in the related and unrelated conditions, to ensure that neither the target nor the distractor generated more looks due to the uncontrolled properties of the words (example in Table 2). These conditions yielded a total of 144 critical items.



Figure 1. Sample trial in semantically related (a) and unrelated (b) conditions, with the target *avión* ‘airplane’, the distractors *vuelo* ‘flight’ (a) and *hueso* ‘bone’ (b), and the fillers *tijera* ‘scissors’ and *plancha* ‘iron’.

Table 2. Sample target stimuli set.

	Related	Unrelated
Word 1	<i>avión—vuelo</i> ‘airplane’—‘flight’	<i>avión—hueso</i> ‘airplane’—‘bone’
Word 2	<i>vuelo—avión</i> ‘flight’—‘airplane’	<i>vuelo—hueso</i> ‘flight’—‘bone’

Semantic relatedness was operationalized using simplified point mutual information (PMI) scores obtained with lightweight metrics of semantic similarity (LMOSS) (Recchia and Jones 2009). This model calculates semantic similarity based on co-occurrence in a trained corpus, in this case, *Corpus del Español* (Davies 2016–) (Supplementary Materials, Supplementary S3). Targets and distractors were matched for frequency and orthographic length based on metrics from the EsPal database (Duchon et al. 2013). Fillers were words with low PMI scores in relation to all the other words in each display. If the target was salient due to the presence of diacritics and/or a complex syllabic structure, then at least one of the other three words in the display had similar characteristics to avoid visual saliency.

An additional 72 filler trials were created for this task. In this case, the words in each display were always semantically unrelated (had low PMI score), and the target/distractor had the same onset syllable (*esposo-estrella* ‘husband-star’) or a different one (*radio-cielo* ‘radio-sky’). Filler trials were included in the experiment to deter participants from finding a pattern in the semantic-related/-unrelated critical trials, but they were not analyzed for this paper.

For each four-word display, a short sentence in Spanish was presented aurally, which included the target word in the second half of the sentence. The beginnings of the sentences were semantically neutral so that any of the four words in the display would fit in that context. In approximately the middle part of the sentence, there was a biasing verb that pointed only to the target or to the target and distractor, in the semantically related condition (*El profesor se acostó en la arena/la playa a leer* ‘The professor **lay down** on the sand/the beach to read’). To avoid participants using intonational contour as a cue, half of the sentences had the target at the end, and the other half did not (*El profesor se acostó en la arena a leer* ‘The professor lay down on the sand to read’ / *Las amigas conversaban mientras se subían al avión* ‘The friends were talking while they boarded the airplane’).

The sentences and the individual words of each display were recorded by a female native Spanish speaker in a sound-attenuated booth using a USBPre2 connected to a Dell desktop computer in Praat (Boersma and Weenink 2017) and with a Shure SM58 microphone and a stand. To avoid phonetically cueing the participants into the target due to coarticulation, the speaker recorded the sentences without the target phrase (noun and preceding article, if any) and in place just produced a bilabial /m/ sound (*El profesor se acostó en /m/ a leer*. ‘The professor lay down on /m/ to read’). Then, the targets were recorded with the corresponding intonation and spliced into the sentence. This was particularly useful when the same sentence had two possible targets, as in the semantically related condition.

To normalize the recorded sentences, the following procedures were followed. First, each sentence was divided into 3–4 sound files, each corresponding to a region of interest for analysis: (1) *preverbal material*, (2) *biasing verb*, (3) *target*, and (4) *posttarget material* (if any). The biasing verb segment included clitics, such as reflexive pronouns. The target segment included the preceding article and prepositions (if any). The reason for including prepositions within this region was mainly prosodic; it would be less naturalistic to splice a pause breaking a prepositional phrase. The intensity of each sound file was normalized to 65 dB, and the boundaries were moved to the nearest zero crossing. Then, the segments of each sentence were concatenated, adding a fixed 500 ms pause before and after the biasing verb (after regions 1 and 2, see Table 3). These pauses were added to give participants enough time to launch anticipatory looks to the target and/or the semantically related distractor, as research shows that it takes approximately 200 ms to plan and launch a saccade (Hallett 1986). Finally, the beginning and final 5 ms of each sentence were faded to eliminate any possible noise inside the booth (e.g., mouse clicks).

Table 3. Sample regions of auditory stimuli for visual world task.

<i>El profesor</i> The professor		<i>se acostó</i> lay down		<i>en la arena</i> on the sand		<i>a leer.</i> to read.	
preverb	500 ms	verb	500 ms	target		posttarget	
REGION 1	PAUSE	REGION 2	PAUSE	REGION 3		REGION 4	

Within the semantically related condition, a potential concern is that the prenominal article may disambiguate between the target and distractor when they have a different gender (e.g., *Al hombre le revisaron el yeso/la muleta* ‘They checked the_{MASC} man’s cast/the_{FEM} man’s crutch’). To circumvent this possibility, the target phrases were recorded without any pauses between the determiner and the noun, as would occur during natural speech. This way, even if the participants noticed the gender difference, there was not enough time to launch anticipatory looks.

A second potential problem is that including fixed pauses before and after the verb may lead listeners to develop a strategy in which they expect to hear the target after the second pause, without attending to the biasing verb or any other part of the sentence. However, there is enough variability in the structure of the sentences to where this strategy may not be useful. In some cases, the target is a noun phrase and, in other cases, a prepositional phrase; some of the filler items have verbs as targets, and, on occasion, the biasing verb was a verb phrase instead of a single word (*El adolescente tenía que leer*(biasing verb) *y*

escribir(target) *para la escuela*. ‘The teenager had to read and write for school’). Therefore, not all the sentences have a basic SVO structure, which would favor specific strategies. Finally, even though there are two fixed 500 ms pauses inserted in the audio, they are not the only pauses present in the sentences. While recording, the speaker included pauses that remained uncontrolled in length to render more naturalistic stimuli.

For the experiment, four lists were created, each with 36 critical and 36 filler items. Half of the critical items in each list were in the semantically related condition and half were in the unrelated condition. The different versions of each critical item (Table 2) were distributed across lists so that participants never saw the same item twice on a list. For example, if the related target–distractor combination *avión-vuelo* ‘airplane-flight’ appeared on list A, then the unrelated version *avión-hueso* ‘airplane-bone’ appeared on list B, while the reverse related combination *vuelo-avión* ‘flight-airplane’ appeared on list C, and the reverse unrelated *vuelo-hueso* ‘flight-bone’, on list D. The filler items were divided in half, and the same fillers appeared in lists A and C or B and D.

2.2.2. Vocabulary Training

Based on the responses of native Spanish speakers to a word association task, 180 Spanish words were selected (Supplementary Materials, Supplementary S4). For each word, an image that pictured the object was created using the platform Story Board That (Story Board That 2022). The purpose of this study was to tap into the English speakers’ L1 semantic network to exploit those rich connections when developing their L2 Spanish network. Thus, to ensure that the images evoked the right concepts for English speakers, the images were normed. Using the crowdsourcing site Prolific (www.prolific.co), 100 native English speakers were recruited and asked to name, in English, the pictures designed for the training. Their responses were rated for accuracy, coding as 1 for a response that matched the intended concept in the picture or 0 if the name given was not the meaning intended. The average accuracy in this norming task was 92.66%.

The Spanish words were recorded in isolation and normalized following the same procedure explained in the visual world section. The same sound files of the words in isolation were used for the visual world task and vocabulary training.

Participants were assigned to one of two training conditions: thematic list (TL) or visual scene (VS). Different stimuli were prepared for each condition.

Thematic list condition—this training condition mimics the vocabulary presentation found in L2 textbooks. Traditionally, words are taught in thematic lists in each chapter (food, professions), with a progression in theme complexity. The 180 words selected were classified into themes, using as a guide the Spanish textbooks *Mosaicos 2* and *Mosaicos 3* (Olivella de Castells et al. 2013, 2015). When a word was not found in these textbooks, other print and online Spanish textbooks were used as a reference to classify each word (Tamariz 2000; Terrell et al. 2001; González-Aguilar and Rosso-O’Laughlin 2005; Potowski et al. 2012; Cubillos 2015; Mir and Bailey de las Heras 2015; Lumen Learning 2018–2019). Then, the themes were organized into an increasing complexity level, using the order from the textbooks *Mosaicos 2* and *Mosaicos 3* (Olivella de Castells et al. 2013, 2015). For example, words like *lápiz* ‘pencil’ and *reloj* ‘clock’ were under the theme ‘classroom items’ in the textbooks, and it is the first topic typically taught at beginner levels. So, they appeared at the beginning of the training. Meanwhile, words such as *vidrio* ‘glass’ and *reciclaje* ‘recycling’ belonged to the theme ‘environment’, and they are usually learned at intermediate-low levels. So, they appeared later in the training.

Visual scene condition—for the visual scene training, six visual scenes were created using the platform Story Board That (Story Board That 2022). Each scene depicted at least two characters in each of these situations: shopping, dinner party, cleaning, sporting event, traveling, and hospital visit. Each scene included 30 vocabulary items for the participants to review, for a total of 180 words as in the list-by-theme training condition. Because of the density of items, each scene was divided into two panels, with 15 items in each one. This

allowed for the objects to be spatially distributed, easier to distinguish, and also to show a progression of a story, similar to a comic strip (Figure 2).



Figure 2. Sample visual scene for training.

The distribution of objects across visual scenes was decided while taking into consideration the experimental trials of the visual world task, to avoid creating artificial associations between semantically unrelated words that could bias the results. To elaborate, words that were semantically related and appeared as target–distractor pairs in the visual world task were trained in the same visual scene and the same panel to train the connection between the items, as would occur in an immersion setting. Additionally, words that were semantically unrelated and appeared as unrelated distractors or fillers in the visual world task were trained in separate scenes from the target. To illustrate, the words *lavadora* ‘washer’ and *secarropa* ‘dryer’ were trained within the same panel in the same visual scene (Figure 2), whereas *aderezo* ‘dressing’ was trained on a different visual scene, since it was presented as an unrelated filler in the same display of the visual world task. In this example, if *aderezo* ‘dressing’ and *lavadora* ‘washer’ had been trained in the same scene and then included in the same visual world trials, this could inadvertently establish an artificial connection between the two items, even though they have a low semantic relatedness score.

2.3. Procedure

Native Spanish speakers completed the study in a single session of approximately 75–90 min. L2 Spanish learners completed the study in four sessions of 75–90 min, each of them approximately one week apart. Figure 3 shows the activities completed in each session.

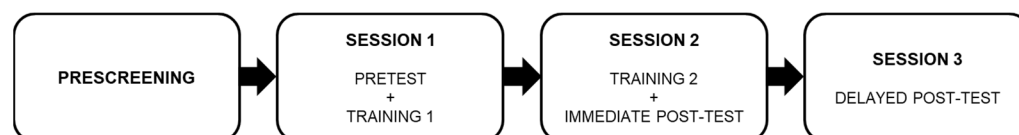


Figure 3. Procedure for L2 Spanish learners.

2.3.1. Prescreening

For the prescreening, participants accessed a Google site that contained the consent form, a link to connect with the experimenter via Zoom (Zoom Video Communications 2022), and the instructions and materials to complete each task. At the beginning of the session, participants were asked to share their computer audio, and the session was recorded on Zoom (Zoom Video Communications 2022) for coding purposes. The tasks were presented

on the Google site, and participants completed each of them while being guided remotely by the experimenter. The order of the tasks was as follows:

- English category fluency;
- Spanish category fluency;
- Digit span;
- Language history questionnaire.

The category fluency task was first completed in the dominant language so that participants would get acquainted with it and perform at their best in the nondominant language.

2.3.2. In-Lab Sessions

The in-lab sessions were approximately 1 week apart, and the task order varied slightly across them. Table 4 summarizes this order, and the procedure of each task is detailed below.

Table 4. Sample regions of auditory stimuli for a visual world task.

	Session 1	Session 2	Session 3
Purpose	Pretest + Training 1	Training 2 + Immediate Post-Test	Delayed Post-Test (No Training)
Task Order	Visual World Training	Training Visual World	Visual World

Training—this task was completed on Experiment Builder (SR Research 2004–2015b) using an MH-5 Millikey button box and Sennheiser HD206 over-the-ear headphones. During this task, participants saw either individual pictures on the screen (TL condition) or visual scenes (VS condition) (Figure 4). In the TL condition, participants saw and heard the name of each object in Spanish (Figure 4a), and they were instructed to repeat it aloud and then press any button in the button box to continue to the next object. The images were organized into thematic lists, following the order found in the textbooks *Mosaicos 2* and *Mosaicos 3* (Olivella de Castells et al. 2013, 2015). In the VS condition, participants saw 6 visual scenes, one at a time, and they were instructed to press any button on the button box, so that one of the objects on the scene was highlighted with a red circle and enlarged in a bubble, as they heard and read the name of the object (in the example, *lavadora* ‘washer’ in the back of the scene is enlarged, Figure 4b). Participants were also instructed to repeat the name aloud before pressing any button on the button box to move on to the next objects. They repeated this procedure for each of the 6 scenes. In both training conditions, the audio recordings were the same used for the isolated words in the visual world task. In each case, participants were trained in 180 words (30 per scene in the VS condition).



Figure 4. Sample training trial for TL condition (a) and VS condition (b).

Visual World—this task was presented on Experiment Builder (SR Research 2004–2015b) using a Dell mouse as an input device and Sennheiser over-the-ear headphones. The participants' eye movements were recorded using an Eyelink 1000 Portable Duo Eye-Tracker (SR Research 2004–2015b), with a head mount. Participants' right eyes were tracked at a sampling rate of 1000 samples per second. A 9-point calibration grid was used. The task was divided into 3 blocks of 24 trials each, and they were preceded by a short practice of 4 trials. The eye tracker was recalibrated between blocks if participants decided to take a break and shift their position. During each trial, participants were presented with 4 Spanish words on the screen, in a diamond-shaped array, using a 40-point Times New Roman white font on a black background. The diamond array was chosen rather than the more classic square array to avoid the corners of the screen, which are often the least precise points during calibration. The trials included a preview for participants to familiarize themselves with the words. During the preview, a green rectangle of 400 W × 350 H pixels appeared around each word, and the participants heard it pronounced with an onset delay of 250 ms. The order in which each word was highlighted and pronounced was always the same: top, left, right, bottom. After the preview, a fixation cross appeared on the screen for 500 ms. Then, the words reappeared in the same position, and the sentence was played through the headphones. Participants were instructed to click on the word mentioned in the sentence, and they were asked to wait until the sentence was done playing. The mouse cursor was a small red triangle on the screen, and it did not appear until the preview period was over (i.e., after the fixation cross, Figure 5). The accuracy of the mouse clicks was recorded, as well as eye movements throughout the entire trial.

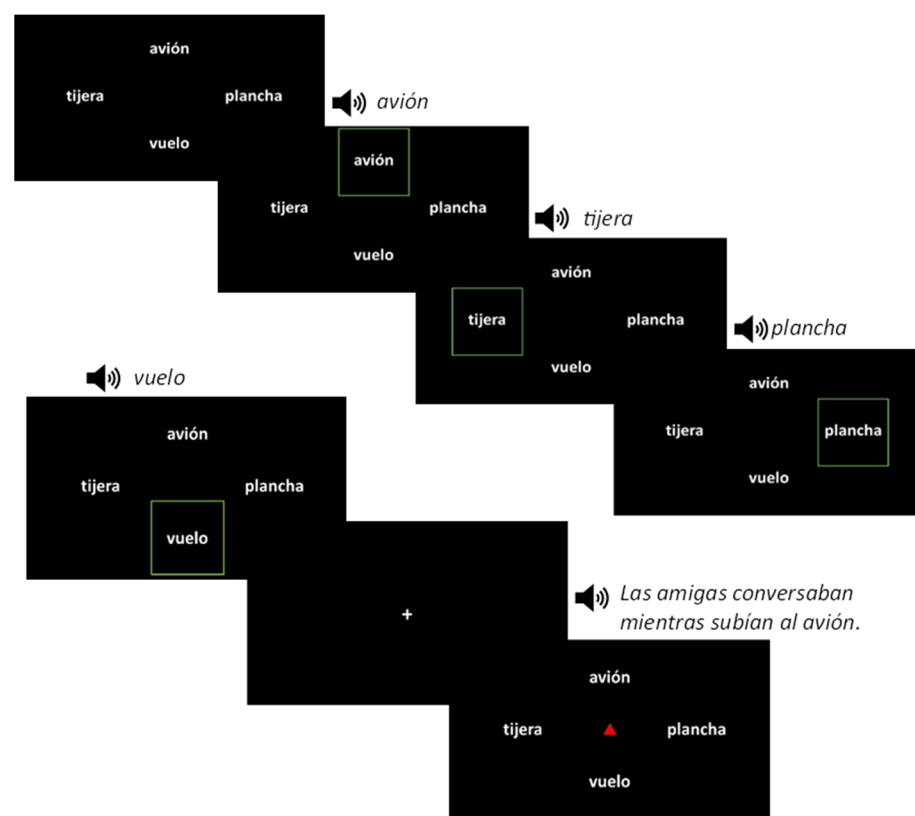


Figure 5. Experimental trial progression for the visual world task.

In each list and block, the position of the words was counterbalanced so that the target, distractor, filler 1, and filler 2 appeared an equal number of times in each of the four locations of the display. Furthermore, the four words did not appear always contiguous in that order (i.e., target, distractor, filler 1, and filler 2 in this order, clockwise).

2.4. Analysis

The eye-tracking data was exported using SR Research Data Viewer software (SR Research 2004–2015a). An interest period was set from the onset of the biasing verb until the offset of the target noun. A time course (binning) report was used to export the data. This report was set to bin time into 10 ms bins; it excluded samples that fell outside of predefined interest areas and samples during blinks or saccades. Trials for which the target object had not been correctly identified or trials that generated no response from the participants were excluded from the eye-movement analyses (0.2% for Session 1, 0.07% for Session 2, and 0.17% for Session 3). All further analyses were conducted in R (R Core Team 2013).

The fixations were time-locked to the onset of the biasing verb preceding the target noun and included a 200 ms baseline for the time it takes to plan and launch a saccade (Hallett 1986). Differential proportions of fixations to the target (DPFT) were then calculated for the analysis by subtracting the proportions of fixations to the distractor from the proportions of fixation to the target. Two analyses were conducted over predetermined interest periods (time windows). The first interest period was from 960 ms to 1460 ms, and it included the 500 ms pause after the biasing verb and before the target onset. This window represents the period where participants had heard the biasing verb, but had not yet heard the critical noun, and is the window necessary to look for effects of semantic interference and anticipatory processing. The second interest period was set from 1460 ms to 2110 ms and included the target (average duration 650 ms).

The DPFT was analyzed with linear mixed-effects models (LME) using the Buildmer (Voeten 2020) and lmerTest (Kuznetsova et al. 2016) packages in R (R Core Team 2013). Two models were conducted for each interest period. The first set of models compared the pretest data of L2 learners (Session 1) with the native Spanish speakers' data, and it included fixed effects of condition (semantically related vs. semantically unrelated) and first language (English vs. Spanish). The semantically related condition and L1 English served as the baseline to which all comparisons were made. The model also included random effects of participant and target item on the intercept. For the participant intercept, the condition was included as a random slope. For the target item, L1 was included as a random slope. L2 proficiency level (operationalized as the category fluency score) was not included as a predictor in any of the models because this study did not intend to test L2 proficiency effects, and participants were recruited from a uniform learner sample. To corroborate, a separate set of linear mixed-effects models was conducted on the L2 data of session 1, with median-centered proficiency as a predictor. Proficiency was not a significant predictor in any model (Supplementary Materials, Supplementary S5).

The second set of models compared the pretest, immediate post-test, and delayed post-test data (Sessions 1, 2, and 3, respectively) of L2 learners across the training groups. These models included fixed effects of condition (semantically related vs. semantically unrelated), training condition (thematic list or TL vs. visual scene or VS), and session (1 vs. 2 vs. 3). The semantically related condition, TL training condition, and Session 1 served as the baseline to which all comparisons were made. The model also included random effects of participant and target item on the intercept. For the participant intercept, we included the condition and session as random slopes. For the target item, the training group and session were included as random slopes. In these models, the Buildmer package (Voeten 2020) takes the maximal model specified by the user (Barr et al. 2013; Bates et al. 2015), and, if it does not converge, it automatically performs a backward stepwise elimination based on the significance of the change in log likelihood.

3. Results

3.1. Comparing Native Speakers and L2 Learners during Pretest

Figure 6 shows the differential proportion of fixations to target (DPFT) split by native language. Time in milliseconds is presented on the x-axis, and the DPFT on the y-axis. In this figure, data points at zero DPFT reflect equal proportion fixations to the target and

distractor, and points above zero reflect that the participants were looking more at the target than the distractor. The regions of interest are highlighted in dark-shaded grey (Verb and Target) and light-shaded grey (Pause) and include a 200-ms baseline for the time it takes to plan and launch a saccade (Hallett 1986). The first region of interest is the pause between the biasing verb and the target. During the pause, we can expect to see differences between the related (black line) and unrelated (red line) conditions, when the listeners have heard the biasing verb (e.g., *subir* ‘board’). At this point, participants would be able to cue into the upcoming target in the unrelated condition (*subir-avión* ‘board-airplane’ vs. *subir-hueso* ‘board-bone’), but not in the related condition (*subir-avión* ‘board-airplane’ vs. *subir-vuelo* ‘board-flight’). In the pause region, before participants hear the target, this effect would be represented by a DPFT significantly above zero in the unrelated condition (red line) and a DPFT hovering over zero in the related condition, with similar looks to the target and the distractor (black line).

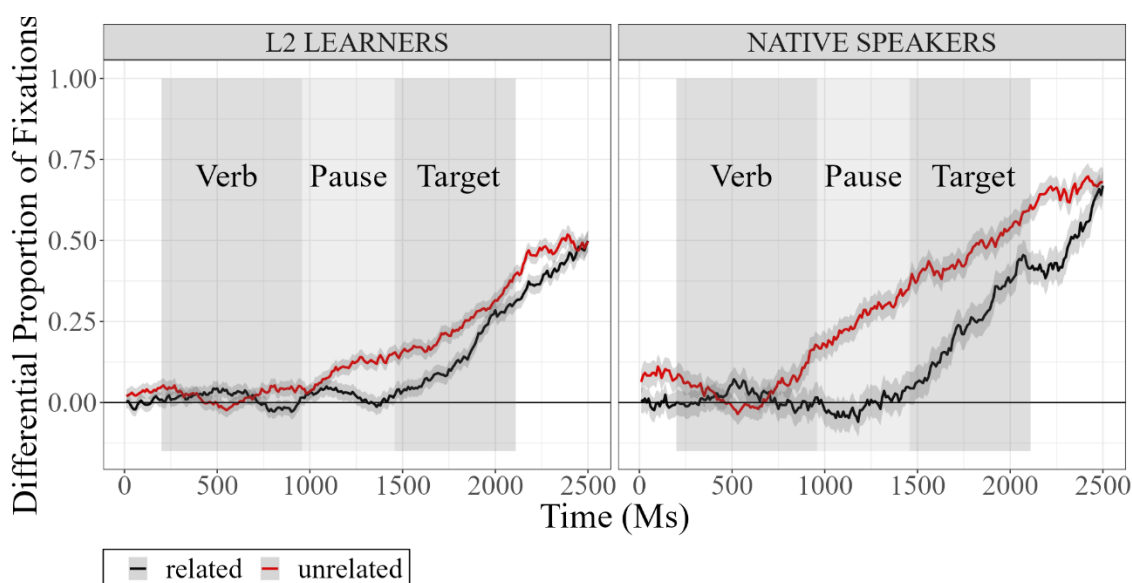


Figure 6. Differential proportion of fixations to target (DPFT) by L1.

The second region of interest includes the target of the sentence. For the related condition, it is expected that the onset of the target would disambiguate between the target and the distractor. At this point, the semantic interference should end, and the black line (DPFT in related condition) would converge with the red line (DPFT in unrelated condition), as in both cases participants are directing their gaze to the target word.

Figure 6 shows that, in the unrelated condition, the native Spanish speakers (right panel) directed their gaze to the target (the only plausible word in the display) while hearing the biasing verb in the unrelated condition (red line). In the related condition (black line), however, the line hovers around zero for most of the duration of the verb and pause, indicating that native speakers were looking equally at the target and the distractor until the target onset. Therefore, native Spanish speakers were aware of the semantic connection between target and distractor (*avión-vuelo* ‘airplane-flight’) and that either one could have followed the biasing verb (*subir* ‘to board’). This competition effect continues into the target region, as shown by the differences between the black and red lines. If the competition could be easily overcome, we would see the black and the red lines quickly converging upon target onset, which is not observed in native speakers, who were still launching some looks to the semantically related distractor after hearing the target.

The L2 learners (left panel) showed a similar pattern to the native speakers in the pause region, but the effect seems weaker and with a later onset than the native speakers. Upon hearing the target, however, L2 learners immediately directed their gaze to it, discarding the semantically related distractor in the related condition. Furthermore, the two solid lines

overlap early on, which may indicate that, for the L2 learners, the semantic competition effect was not as long-lasting as for the native speakers. These results may indicate that the semantic link between the related target and distractor is not as strong for L2 learners as it is for native speakers. Note, however, that the difference between the native speakers and L2 learners seems to be mostly driven by the unrelated condition (red line) and not the related one (black line). Native speakers seem to commit much earlier to a target in the semantically unrelated condition compared to the L2 learners. This and other potential reasons for these differences are explored in the discussion.

Table 5 presents the results of the linear mixed-effects model (LME) with the best fit on the differential proportion fixations to target in the pause region during the pretest. As summarized in Table 5, the nonsignificant estimate for the intercept suggests that the DPFT for the target items in the related condition was not statistically different from zero in L2 learners, indicating that they were looking at the target and distractor equally during the pause in the related condition. From the remaining estimates, we can draw three main conclusions.

Table 5. Results of LME on pretest for the pause region.

	Estimate	Std. Error	t Value	Pr (> t)
Intercept	0.02	0.02	0.84	n.s.
Condition	0.08	0.02	3.24	<0.01
L1	−0.03	0.05	−0.71	n.s.
Condition: L1	0.19	0.05	3.81	<0.001

First, the significant estimate for a condition with a positive slope shows that L2 learners' DPFTs were significantly higher for the unrelated condition than for the related condition in the pause region. Second, the negative slope of the main effect of L1 is an indicator that native speakers were looking at the target less than the L2 learners during the pause in the related condition, but this effect was not significant. Finally, the significant two-way interaction between condition and L1 confirms that the difference between related and unrelated conditions was much larger for native speakers than for L2 learners in the pause region. This significant interaction may be an indicator that the semantic connections between words in the lexicon are stronger for native speakers, yielding an effect of condition, whereas L2 learners showed a weaker effect. Another possibility is that this interaction is the product of a timing difference since L2 learners showed a split between the conditions later than the native speakers. Both possibilities are in line with typical L2-learner behavior, in which some effects tend to appear later and be weaker when compared to native speakers, who have had extensive immersive exposure to the language, as seen, for example, in the anticipatory processing literature (Kaan 2014).

Some of these effects carry over into the target region. Table 6 presents the results of the linear mixed-effects model (LME) with the best fit on the differential proportion fixations to target in the target region during the pretest. As summarized in Table 6, the significant positive estimate for the intercept indicates that the DPFTs for the target items in the related condition were statistically different from zero in L2 learners, meaning that they were looking at the target more upon hearing it in the related condition.

Table 6. Results of LME on pretest for the target region.

	Estimate	Std. Error	t Value	Pr (> t)
Intercept	0.15	0.03	4.52	<0.001
Condition	0.09	0.03	3.57	<0.001
L1	0.10	0.05	2.03	<0.05
Condition: L1	0.12	0.05	2.37	<0.05

The main effect of L1 indicates that the native Spanish speakers were looking at the target more than the L2 learners in the related condition. The main effect of the condition, similar to the pause window, shows that L2 learners directed their gaze toward the target significantly more in the unrelated condition than in the related condition. Finally, the significant two-way interaction between L1 and condition shows that the difference in DPFT between the unrelated and related conditions was larger for native Spanish speakers than for L2 learners, an effect already observed in the pause region.

3.2. Comparing L2 across Sessions

Figure 7 shows the differential proportion of fixations to target (DPFT) split by session (pretest, immediate post-test, and delayed post-test) and training group (TL and VS). Time in milliseconds is presented on the x-axis, and the DPFT on the y-axis. In this figure, data points at zero reflect the equal proportion of fixations to the target and distractor, and points above zero reflect that the participants were looking more at the target than the distractor. The regions of interest are highlighted in dark-shaded grey (verb and target) and light-shaded grey (pause) and include a 200 ms baseline for the time it takes to plan and launch a saccade (Hallett 1986). The top row of Figure 7 shows the data for each session for the TL group, and the bottom row shows the progression of the VS group. The three columns correspond to the three sessions.

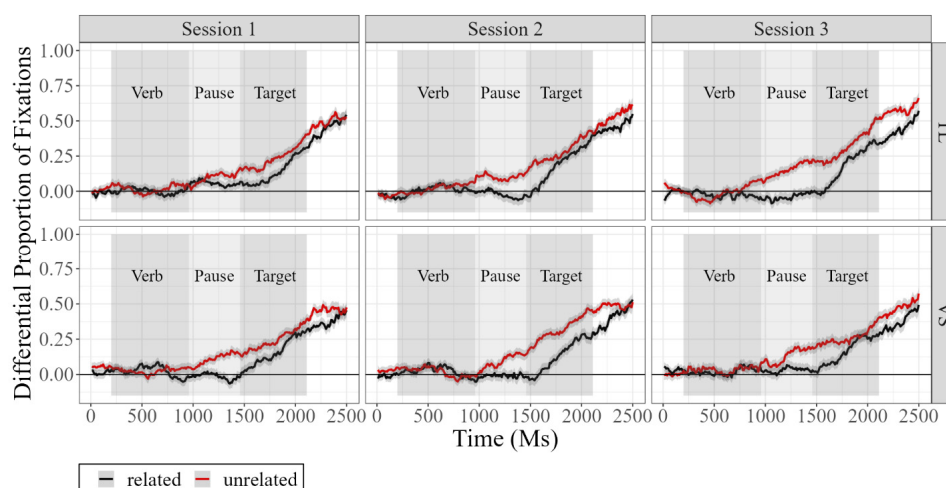


Figure 7. Differential proportion of fixations to target (DPFT) by session and training.

In the pause region, the TL group shows changes across each testing session on the related (black line) and unrelated (red line) conditions. The DPFT gets closer to zero from Session 1 to Session 2 in the TL group in the related condition, which means a similar proportion of looks to the target and the distractor, and, hence, semantic competition, which appears to increase with training. In the unrelated condition, the TL group shows a higher DPFT, especially in the last session during the pause window. The VS group, however, shows more consistency across sessions in the pause region, which may be an indicator of fewer or no changes in the training.

In the target window, there are other differences between the two training groups. In Session 1, in both groups, the DPFT in the related (black line) and unrelated (red line) conditions almost converge upon hearing the target. This means that any competition effect that could arise from the related distractor quickly disappears upon hearing the target. In Session 2, however, there are noticeable differences between the two groups. The TL group shows a similar pattern as in Session 1, with the two lines quickly converging upon hearing the target. The VS group, however, looked at the target significantly more in the unrelated condition than in the related condition in the target window. This pattern is similar to the behavior of native speakers (see Figure 6 above), for whom the semantic competition effect seemed to last longer in the target region, possibly due to the strength

of connections between the target and the semantically related distractor. In Session 3, however, this difference between the two groups is attenuated, as they exhibit a similar pattern of fixations.

Table 7 presents the results of the linear mixed-effects model (LME) with the best fit on the differential proportion fixations to target in the pause region across sessions and training groups. As summarized in Table 7, the nonsignificant estimate for the intercept indicates that the DPFTs for the target items in the related condition were not statistically different from zero in the TL group on Session 1 because they were looking at the target and distractor equally during the pause in the related condition.

Table 7. Results of LME across sessions and training for the pause region.

	Estimate	Std. Error	t Value	Pr (> t)
Intercept	0.04	0.03	1.23	n.s.
Condition	0.07	0.03	2.13	<0.05
Session 2	−0.07	0.03	−1.89	n.s.
Session 3	−0.09	0.04	−2.29	<0.05
Training	−0.03	0.04	−0.81	n.s.
Condition: Session 2	0.07	0.01	6.45	<0.001
Condition: Session 3	0.01	0.01	13.69	<0.001
Session 2: Training	0.03	0.04	0.84	n.s.
Session 3: Training	0.12	0.04	2.67	<0.05
Condition: Training	0.04	0.04	1.00	n.s.
Condition:Session 2: Training	−0.09	0.02	−5.94	<0.001
Condition:Session 3: Training	−0.16	0.02	−10.58	<0.001

The main effect of the condition was significant with a positive slope for the TL group, suggesting that the DPFTs were higher in the unrelated condition compared to the related condition during the pause. There was a significant main effect for Session 3 with a negative slope, indicating that during the delayed post-tests the TL group was looking at the target less in the related condition, exhibiting more competition with the semantic distractor compared to the pretest. This difference was not significant for Session 2 (immediate post-test). Indeed, in Figure 7, the black line is hovering over zero during the delayed post-test, whereas, in the pretest, the line is slightly above zero for the TL group. Finally, the main effect of training shows that there was no significant difference between the TL and the VS groups in Session 1. This was expected since the participants for each group had been matched for proficiency and working memory, and they had received no training at the time of the pretest.

In addition to the main effects, the model on the pause window yielded some significant two-way and three-way interactions. First, the significant interactions between condition and session (2 and 3) indicate that, for the TL group, the difference between the unrelated and related conditions increases significantly in Sessions 2 and 3. Therefore, after training, the TL group has a significantly higher DPFT in the unrelated condition during the pause (red line in Figure 7) compared to the related condition (black line). Second, the significant interactions between Session 3 and the training show that the decrease in DPFT in the related condition in the delayed post-test compared to the pretests occurred for the TL group, but not for the VS group.

Finally, the three-way interactions between condition, session, and training group show that the incremental difference between related and unrelated conditions in the post-tests compared to the pretest experienced by the TL participants does not occur for VS participants in the pause region. These interactions may be related to some marginal differences observed between the groups in the main effect of training, where the VS group appeared to show semantic competition between the target and the distractor in the related condition from the pretest, whereas the TL group did not. These interactions may not be significant with a larger sample size.

The differences observed between the groups and across sessions are more nuanced for the target region. Table 8 presents the results of the linear mixed-effects model (LME)

with the best fit on the differential proportion fixations to target in the target region across sessions and training groups. As summarized in Table 8, the significant positive estimate for the intercept indicates that the DPFTs for the target items in the related condition were statistically different from zero in the TL group in Session 1, as they were looking at the target significantly more than the distractor upon hearing the target in the related condition.

Table 8. Results of LME across sessions and training for the target region.

	Estimate	Std. Error	t Value	Pr (> t)
Intercept	0.14	0.04	3.78	<0.001
Condition	0.10	0.02	3.99	<0.001
Session 2	0.04	0.04	1.12	n.s.
Session 3	0.03	0.03	0.91	n.s.
Training	0.03	0.04	0.68	n.s.
Condition: Session 2	0.01	0.01	1.20	n.s.
Condition: Session 3	0.06	0.01	7.07	<0.001
Session 2: Training	−0.08	0.04	−1.72	n.s.
Session 3: Training	−0.03	0.04	−0.86	n.s.
Condition: Training	−0.02	0.03	−0.62	n.s.
Condition: Session 2: Training	0.12	0.01	9.32	<0.001
Condition: Session 3: Training	−0.03	0.01	−2.55	<0.05

The main effect of the condition in the target window reveals that the TL participants had higher DPFT in the unrelated condition than in the related condition. This difference in Session 1 is consistent across groups, as indicated by the nonsignificant main effect of the training group in the target window. The similarity across groups was expected in Session 1 (pretest) before the participants had any training.

In addition to the main effect of the condition, there were two significant interactions in this model. The significant two-way interaction of the condition and Session 3 (delayed post-test) shows that the difference between related and unrelated conditions increased compared to the pretest in the TL group. This difference, however, was not significant from Session 1 to Session 2 (immediate post-test). Finally, the significant three-way interaction of condition, Session 2, and training group captures the data observed in the middle panel, bottom row of Figure 7. This interaction suggests that the difference between DPFT in the related and unrelated conditions becomes larger in the immediate post-test (Session 2) for the VS group only. This difference, however, seems to decrease in Session 3, which may be an indicator of a short-lived effect of training. Alternatively, this may be a practice effect, due to the repetition of the task. The VS Session 2 panel referenced resembles the native-speaker panel for Session 1 (Figure 6). Figure 8 shows the DPFT of native speakers (naturalistic exposure, NE) and the two training groups in Session 2. To corroborate this similarity between NE and VS, a post hoc analysis was conducted.

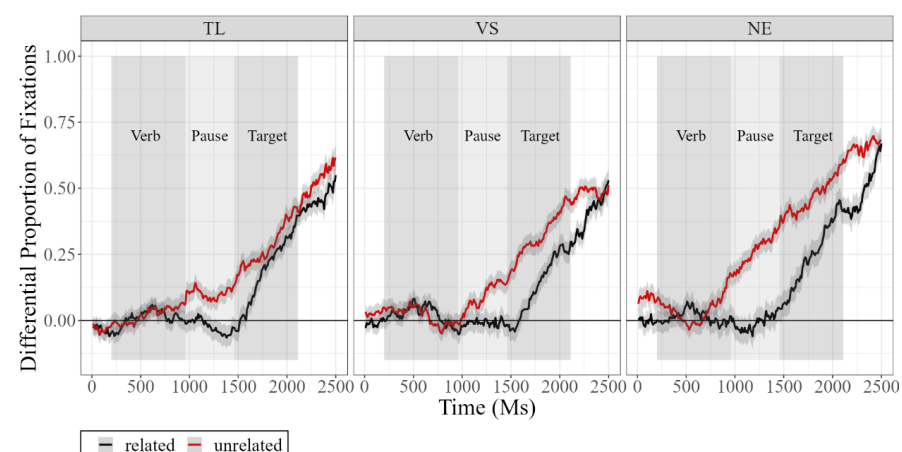


Figure 8. Differential proportion of fixations to target (DPFT) by exposure type.

An additional linear mixed-effects model was carried out by comparing the DPFTs in three exposure groups in Session 2 (immediate post-test). The three groups refer to the TL group (exposed to thematic lists), the VS group (exposed to visual scenes), and the native-speaker group (naturalistic exposure, NE). The model included condition (related vs. unrelated) and exposure group (TL vs. VS vs. NE) as fixed effects. The related condition and NE group served as the baseline against which all comparisons were made. The random effects of participant and target item were also included in the intercept. For the participant intercept, the condition as a random slope was included. For the target item, the exposure group was included as a random slope.

Table 9 summarizes the results of the post hoc linear mixed-effects model (LME) with the best fit on the differential proportion fixations to target in the target region across exposure groups. The significant estimate of the intercept suggests that the NE group was looking at the target more than the distractor in the target region, which is unsurprising. From the remaining main effects and interactions, some conclusions can be derived.

Table 9. Results of LME after treatment for the target region.

	Estimate	Std. Error	t Value	Pr (> t)
Intercept	0.25	0.05	5.39	<0.001
Condition	0.21	0.04	5.45	<0.001
TL Exposure	−0.06	0.06	−1.04	n.s.
VS Exposure	−0.11	0.05	−2.10	<0.05
Condition: TL Exposure	−0.12	0.05	−2.36	<0.05
Condition: VS Exposure	−0.02	0.05	−0.47	n.s.

The main effect of the condition confirms the result that the NE group had a significantly higher DPFT in the unrelated condition than in the related condition in the target region, as previously discussed. The main effect of VS exposure indicates that the VS group has significantly lower DPFT in the target region and the related condition compared to the NE group, but this is not the case for TL. The most meaningful comparisons that confirm the similarity between the NE and VS groups are the two-way interactions between condition and exposure. The significant difference between the related and unrelated conditions found in the NE group is also present in the VS group, as suggested by the nonsignificant interaction between the condition and the VS. For the TL group, however, the difference between the two conditions is significantly smaller than in the NE group.

This similarity between the NE and the VS groups after exposure suggests that the visual scene training can approximate the naturalistic exposure experienced by native speakers when learning vocabulary. If this is the case, using visual scenes may enhance instruction in the L2 classroom, emulating an immersion experience and strengthening the connections in the lexicon. There is a caveat to this finding, however. The similarities between NE and VS seem to be driven by changes in the unrelated condition, not the related one, so this may be a limitation when making claims about the strength of semantic connections.

In sum, the data yielded three main findings:

- (1) Native speakers and L2 learners were significantly different at pretest, both in the pause and target regions. Both groups showed more looks to the target in the unrelated condition compared to the related condition, but the magnitude and timing of this effect were significantly different between the two groups;
- (2) The L2 learner groups (thematic list training and visual scene training) developed different behaviors after two rounds of training, with the VS group resembling the native speakers more than the TL group in Session 2 (immediate post-test). These differences across TL and VS, however, decreased by Session 3 (delayed post-test) because the VS group's improvement diminished in the delayed post-test;

- (3) The differences between the two groups in Session 2 seem to be driven by the unrelated trials, rather than the related ones, contrary to what was targeted by the manipulations in the experimental design.

Possible explanations and the implications of each result are further explored in the Discussion.

4. Discussion

This experiment tested how the semantic connections in the L2 lexicon affect real-time comprehension, and whether different types of exposure to vocabulary may strengthen these connections and enhance communication in the L2. Two groups of L2 Spanish learners completed a three-session training study in which they reviewed vocabulary in Spanish under one of two conditions, thematic lists, as found in textbooks (TL), and visual scenes with complex semantic connections, based on co-occurrence (VS). Their semantic connections were assessed with an eye-tracking visual world paradigm, which was repeated in each of the three sessions, serving as pretest, immediate post-test, and delayed post-test, respectively. Each trial in the visual world task included a four-word display, with a target and a distractor that was semantically related (*avión-vuelo* 'airplane-flight') or unrelated (*avión-hueso* 'airplane-bone'), in addition to two fillers. A group of native Spanish speakers completed the visual world task once and served as a baseline of participants with naturalistic exposure.

The main results of this study suggest that (1) native speakers were significantly different than L2 learners during the pretest, with native speakers showing stronger anticipatory effects; (2) the VS group behaved more similarly to native speakers than the TL group during an immediate post-test; and (3) the differences between training groups were driven by the semantically unrelated condition.

First, participants with natural exposure (native speakers) launched significantly more saccades to the target in the unrelated condition compared to the related condition before the target onset. This difference between the two conditions continued even after the target had been mentioned. A similar difference in looks to target between the unrelated and related conditions was observed for the L2 learners, but much smaller compared to native speakers, and the two conditions converged after target onset. In other words, the biasing verb (e.g., *subir* 'to board'), cued native speakers immediately to the target in the unrelated condition (*avión-hueso* 'airplane-bone'), but in the related condition (*avión-vuelo* 'airplane-flight'), they had to wait until target onset. Interestingly, upon hearing the target, native speakers still launched saccades to the semantically related distractor, indicating that they were not able to immediately overcome the semantic competition, even after having enough information to commit to the target. This finding is in line with previous eye-tracking studies that focused on lexical semantics and semantic overlap. Huettig and Altmann (2005), for example, observed that, when participants were presented with semantically related items on a display in a visual world paradigm (e.g., trumpet-piano), the semantically related distractor was favored over unrelated distractors, even after the target was mentioned.

During the pretest, participants with classroom experience (L2 learners before training), showed weaker differences between the two conditions. They were also able to identify the target in the unrelated condition before it was mentioned, but this effect occurred later and was smaller than in the native speakers. Additionally, after hearing the target, L2 learners still launched more looks to the target in the unrelated condition than in the related condition (similar to native speakers), but the difference disappeared quickly into the target region. This behavior may be an indicator that, unlike participants with naturalistic exposure, L2 classroom learners have weaker links between the two semantically related words, and they were able to overcome the competition effect faster than the native speakers. If this is the case, weaker links may result from the way vocabulary is learned during classroom exposure.

This scenario constitutes one possibility, but part of the results challenges this explanation. The difference between the related and unrelated conditions across native speakers and L2 learners seems to be driven by the unrelated condition, not the related one. In other words, both groups exhibit similar behaviors in the related condition, although the semantic interference may be weaker in L2 learners. But, the true difference is in the unrelated condition. Native speakers committed to the target early when they heard the biasing verb and there was no other plausible ending to that sentence. L2 learners, however, seemed to deploy a wait-and-see strategy before fully committing to the target in the unrelated condition.

The response of native speakers to the unrelated condition (launching saccades to the target before hearing it) is unsurprising. Anticipatory looks at the target have been well documented in the eye-tracking literature. When presented with adequate lexical or morphosyntactic cues in a favorable context, native speakers are able to anticipate an upcoming target even before it is mentioned (Altmann and Kamide 1999; DeLong et al. 2005; Dussias et al. 2013; Kaan 2014; Lew-Williams and Fernald 2007, but see Huettig 2015; Huettig and Mani 2016). In the present project, the introduction of a biasing verb (e.g., *subir* ‘to board’) when there was only one possible target (e.g., *avión-hueso-tijera-plancha* ‘airplane-bone-scissors-iron’) constituted a highly restrictive semantic context that favored the generation of expectations. The literature on L2 learners’ anticipatory processing, however, is more nuanced.

Research on anticipation in L2 learners has yielded mixed results (Lew-Williams and Fernald 2010; Grüter et al. 2012; Dussias et al. 2013; Hopp 2013; Dijkgraaf et al. 2017, 2019), and anticipatory processing may be modulated by L2 proficiency (Connell et al. 2021). Additionally, learners tend to engage in anticipatory processing mainly when presented with a reliable cue that triggers a strong representation; otherwise, the cost-benefit of anticipation is too high to engage (Kaan 2014; Kuperberg and Jaeger 2016; Kaan and Grüter 2021).

Taken together, this information may explain the differences observed between the native speakers and the L2 learners during the pretest, mainly in the unrelated condition. Native Spanish speakers, on the one hand, showed strong semantic interference in the related condition (*avión-vuelo* ‘airplane-flight’), and anticipation of the target in the unrelated condition (*avión-hueso* ‘airplane-bone’) upon hearing the biasing verb. These two phenomena combined account for the divergence of the DPFT during the verb, the strong difference between the two conditions in the pause region, and the spillover of this difference into the target for native speakers. L2 learners, on the other hand, before they had any training, showed a similar semantic interference in the related condition, but significantly less looks to the target in the unrelated condition compared to the native speakers. This initial difference begs the question of how the L2 learners may change their behavior over time with training, which leads us to the second main finding.

At the pretest, the two groups were expected to be similar, since they had not had any training and the participants had been matched for L2 proficiency (category fluency score) and working memory (digit span score), and this was indeed the case. The two groups (thematic list training TL and visual scene training VS) were not significantly different at the pretest in the related condition, at the pause region, or the target region, as hypothesized before training.

When the L2 learners were tested in Session 2 (immediate post-test), after two rounds of training, differences began to emerge in the target region between the TL and the VS groups. In Session 2, the VS group maintained a significant difference between DPFT in the unrelated and related conditions in the target region. This means that, even after the target had been mentioned, VS participants overall had a higher proportion of looks to target in the unrelated condition, but, in the related condition, they still launched some looks to the semantically related distractor. This pattern was comparable to the participants with naturalistic exposure (native speakers), who also showed this difference between the two conditions even in the target region. Such a difference was not observed in the TL group, whose DPFT in both conditions overlapped in the target region.

The TL and the VS training differed in the way vocabulary is presented and stored in the learners' lexicon. Note that the vocabulary trained belonged to beginner and low-intermediate Spanish levels, and the learners in this experiment were intermediate. So, they were presumably familiar with the words already. This was intentional because the training did not aim to teach new words but to train connections between words to enhance the L2 lexicon. If the training worked as expected, then the findings would inform L2 instruction and carry over to lower levels where this vocabulary is first introduced. On the one side, the TL group was exposed to vocabulary in isolation organized by thematic lists as would be encountered in a textbook. Hence, this training reinforced the exposure with which classroom-based learners were already familiar, and changes over time were not expected in these learners other than possible practice effects. On the other side, the VS group reviewed vocabulary in visual scenes that established more complex connections between words based on co-occurrence. Because the scenes were built with co-occurrence in mind, sometimes, words did belong to the same theme, as would be found in a textbook (e.g., *árbitro-jugador* 'referee-player' within sports vocabulary or *suéter-falda* 'sweater-skirt' within clothing vocabulary). But crucially, at other times, words with high co-occurrence, as trained on the visual scenes, appear on different lists and even chapters of the same textbook. For example, the word *sol* 'sun' is taught with weather vocabulary (early in beginner courses), *luna* 'moon' with nature vocabulary (late in low-intermediate courses), and *basura* 'trash' is taught with house chores (beginner) and *reciclaje* 'recycling' is taught with environment vocabulary (low intermediate). However, the data from *Corpus del Español* (Davies 2016–) confirms that *sol-luna* 'sun-moon' and *basura-reciclaje* 'trash-recycling' are words with high co-occurrence in everyday speech (see Supplementary Materials, Supplementary S3, for how the co-occurrence scores were computed). Therefore, participants in the VS condition were expected to establish meaningful connections in their L2 lexicon, thus resembling participants with naturalistic exposure in their processing during the visual world task. The difference between the TL group and the VS group would be measured by a large difference in VS between the related (*avión-vuelo* 'airplane-flight') and unrelated (*avión-hueso* 'airplane-bone') conditions, caused by an increase in semantic competition in the related condition due to training when compared to the TL group. This difference was expected to appear in the pause region (upon hearing the biasing verb and before target onset), and potentially spill over to the target region, as was observed in the native speakers.

The present results partially supported these hypotheses, with an unexpected twist. The VS group did indeed resemble the natural exposure group (native speakers) in the immediate post-test (after two training sessions). This was confirmed through the post hoc comparison of the TL and VS Session 2 data with the native-speaker baseline data. In the target region, the VS group showed significant changes in the difference between the two conditions not displayed by the TL group. In this sense, the VS training seemed to have an effect on the L2 learners that was closer to a naturalistic exposure based on the co-occurrence distribution of vocabulary compared to the thematic lists in textbooks. There is a caveat, however; the L2 learners showed from the first session a similar behavior to the native speakers in the related condition, and the changes seemed to be associated with the unrelated condition. This was the third main finding of this experiment, and the last one discussed in this paper.

The lack of major changes in the related condition across time is probably due to learners already experiencing semantic competition before they started the training. Recall the first linear mixed-effects model reported (comparison of native speakers and L2 learners at pretest, Table 5), which included L2 learners and a related condition as the baseline in the pause window. The intercept for this model (DPFT in a related condition for L2 learners) was not significantly different from zero, meaning that L2 learners in the related condition were looking at the target (*avión* 'airplane') and semantically related distractor (*vuelo* 'flight') upon hearing the biasing verb (*subir* 'to board'). This is evidence of a

semantic competition effect taking place even before training occurred. When considering the population included in the sample, this is not entirely surprising.

The participants in this experiment were intermediate L2 learners who were familiar with most of the vocabulary included in the visual world task. Even though these were classroom-based learners who were probably exposed to thematic vocabulary lists, their learning had been taking place with a fully developed L1 lexicon already present. Therefore, it is likely that these learners were proficient enough to use the connections already present in their L1 semantic network to establish associations between L2 concepts. This is even more plausible considering that most of the critical items were concrete nouns, allowing for a high degree of conceptual overlap between the L1 and the L2 (Marian and Kaushanskaya 2007; but see Malt 2010 for counterevidence). The selection of concrete nouns for this experiment was intentional so as to have highly picturable objects that could be included in an image-based training. Unfortunately, this design is limited by a lack of comparison between concrete and abstract concepts, the latter presumably being more distinct between the L1 and the L2. Future research would benefit from incorporating concrete and abstract nouns into training to fully test this hypothesis. Another alternative direction for this line of research should include early learners who are being exposed to these new words for the first time in different training conditions. Ideally, a four-way comparison, including concrete and abstract word training, and novice and intermediate L2 learners, would corroborate these potential explanations of the results.

Now, if both groups of L2 learners were showing a semantic competition effect from the pretest, then differences across groups would not be expected. However, this is not what was found. The VS group and the TL group are significantly different in Session 2 (immediate post-test) due to changes in the unrelated condition. The unrelated condition can be considered a favorable context to generate expectations during speech processing. Indeed, native speakers, who had received naturalistic exposure in Spanish, launched saccades toward the target before it was mentioned in the unrelated condition. L2 learners at the pretest were adopting a wait-and-see approach in the unrelated condition, which was much closer to the related condition compared to the native speakers. After two rounds of training, however, the VS group displayed a similar pattern to the native speakers in Session 2, but the TL group did not. There are a few possible explanations for this difference between the groups across the pretest and immediate post-test.

One unlikely possibility is that L2 learners are experiencing a practice effect due to the repetition of the task, and they were able to tune into the target as soon as they heard the verb in the unrelated condition. The reason why this explanation is not plausible is that, if this were the case, then we would see the same changes across the TL and the VS groups, and that does not happen. Additionally, even though the visual world task was the same, participants completed different lists each time and never saw the same display in two consecutive sessions.

A second possibility relates to the circumstances and conditions that favor the generation of expectations as a useful strategy for L2 comprehension. Kaan (2014) has suggested that anticipation is possible when there is a strong mental representation (see also Kaan et al. 2010; Kaan and Grüter 2021). This means that, to use a morphosyntactic or lexical semantic cue to generate expectations, L2 learners must have a well-established representation of the morphosyntactic or lexical semantic form in the target language; the cost-benefit ratio of generating expectations based on weak representations may be too great to be worthwhile. Therefore, it is possible that the L2 learners began the experiment with fuzzy semantic representations of the words tested, and, then, the VS group developed stronger representations allowing them to generate expectations during the post-test. This explanation poses the question of why the VS training strengthened representations immediately after training, whereas the TL training did not.

Previous research on L2 learning suggests that being exposed to complex word associations is beneficial for vocabulary processing (Finkbeiner and Nicol 2003; Bolger and Zapata 2011). Bolger and Zapata (2011) discuss that contextualized vocabulary presentation promotes depth of processing (Craik and Lockhart 1972; Craik and Tulving 1975), which refers to high levels of semantic engagement. In this sense, it is possible that learning words in complex contexts in the current experiment may have strengthened the semantic representations in the VS group, hence the differences observed in Session 2 between the two training groups.

The VS and the TL training simultaneously compared vocabulary presented in isolation with vocabulary in context and thematic relations with co-occurrence relations. Therefore, this design does not allow for disentangling the effect of isolation/context and theme/co-occurrence, and they must be interpreted together. In this sense, the advantage observed in the VS group can be due to vocabulary being presented in context, thus favoring deep semantic processing. Or, it may be due to the presence of co-occurrences that strengthened connections in the lexicon or a combination of both. Training in co-occurrences may be beneficial to refine the meanings of existing concepts in the lexicon, thus promoting vocabulary depth (Nagy and Herman 1987). Further research should implement experimental designs that can tease apart these factors. Regardless of the mechanism behind the improvement, the VS training seems to approximate a more naturalistic exposure, as experienced by native speakers, in which words are learned in situational contexts where co-occurrence distributions are highly relevant. In sum, the stronger representations built by the VS group may be responsible for their increased anticipatory processing observed in the unrelated condition of the visual world task. This has implications for L2 pedagogy and the development of ecologically valid teaching materials for the classroom, which will be discussed in the next subsection.

Implications for L2 Instruction: Bringing Immersion to the Classroom

Research on L2 acquisition has shown that immersion is highly beneficial for vocabulary learning and fluency gains (Milton and Meara 1995; Llanes and Muñoz 2009; Pizziconi 2017). These benefits are probably due to vocabulary being learned in meaningful contexts and co-occurrence with related words and concepts that favor the integration of new words into the lexicon. The best way to improve the L2 learning experience, then, would be to ensure that learners have the opportunity to become immersed in the L2 language and culture. Of course, this is not always feasible for practical and, more importantly, economic reasons. Although studying abroad is a valuable experience, its costs can be prohibitive for most students in the L2 classroom. Therefore, if learners cannot experience immersion for themselves, can we as L2 educators bring immersion to the classroom? In theory, we can; but first, we need to understand what aspects of immersion make the L2 learning experience so fruitful.

One of the purposes of this project was to begin disentangling which cognitive mechanisms behind immersion may enhance L2 acquisition, with a focus on vocabulary. The findings of this project suggest a few reasons why immersion is so beneficial for L2 vocabulary learning. The first option is that, during immersion, vocabulary is encountered in context, embedded in meaningful situations that are culturally relevant, instead of in isolation or in lists. The usefulness of context in vocabulary learning has been already highlighted in previous work, especially when it includes words from different themes rather than focusing on one at a time (Finkbeiner and Nicol 2003; Bolger and Zapata 2011). For example, Bolger and Zapata (2011) observed advantages in learning novel words embedded in a story when they combined vocabulary related to animals, utensils, body parts, and other themes, compared to stories that only included one theme, such as animals. Some of the traditional L2 teaching methodologies have focused on teaching vocabulary in thematic lists. This practice is rooted in outdated research on monolingual speakers, showing that thematic word lists are easier to memorize than unrelated lists (Bousfield 1953; Cofer 1966; Cohen 1963). However, contextualizing vocabulary offers an opportunity

to create meaningful materials that incorporate stories or situations that are relevant and interesting to students. Additionally, having to process vocabulary in conjunction with other words encourages students to pay more attention to the new words and connect them with other vocabulary in their lexicon, which may be more semantically engaging than merely memorizing lists and their L1 translation equivalents. Learning new words in relation to other vocabulary can help generate richer semantic connections that strengthen the representations of each word in the lexicon (Nagy and Herman 1987), making them less fuzzy and more precise (Bordag et al. 2022).

Therefore, a first step towards incorporating an immersive experience in the classroom would be to present vocabulary in context, which can be textual, audiovisual, or in interactive materials. Some research has studied the role of virtual reality as a way to engage students with new vocabulary in an open-world environment that emulates a study-abroad experience (e.g., Hsiao et al. 2017). Results from this line of research are promising, but the outcome depends largely on an individual student's ability to generate meaningful semantic connections beyond what they have learned in textbooks. Therefore, although some open-ended experiences are valuable, they should be complemented with teaching materials that help all students, including those who are not able to make different word associations on their own. Beyond students' individual differences, there are some limitations to incorporating technology in the classroom, especially for contexts that do not have the socioeconomic resources necessary to set up, for example, a virtual reality lab or even a computer lab. Written (and to an extent audiovisual) materials remain the best alternative to incorporate context into vocabulary learning, as they are readily available in most learning contexts.

The next question that arises is, if we decide to present vocabulary in context, what other words do we use to create such contexts and materials? As already mentioned, context is even more useful for L2 vocabulary learning when it incorporates words from different themes, not just thematic lists converted to text (Finkbeiner and Nicol 2003; Bolger and Zapata 2011). One of the alternatives would be to use authentic materials in the classroom, a practice that has been widely studied and incorporated into L2 learning. The use of authentic materials can be challenging but also highly beneficial for L2 outcomes, as they help increase motivation and engagement in students (for a review see Gilmore 2007).

The use of context and authentic materials are not new ideas. The real problem, and this is where this paper may help, is using better practices when materials need to be artificially created or adapted from authentic sources to ensure they are effective. The creation of L2 teaching materials, then, should be based on the co-occurrence distribution of vocabulary, and this information can be obtained through the study of existing corpora or word-association norms in the target language. Furthermore, the metrics of semantic similarity and co-occurrence used in the present project can be extrapolated to the teaching of other second and foreign languages, not just Spanish. This way, educators can start bridging the gap between students who have the possibility to study abroad or seek other forms of immersion, and those learners who can only access formal instruction in the L2 classroom. Finally, this type of research can start an open conversation between L2 researchers and educators to inform better classroom practices and also bring the voices of educators when assessing the feasibility of incorporating certain changes into curriculum design.

5. Conclusions

The present study tested whether different types of exposure to L2 Spanish vocabulary would affect L2 learners' processing during comprehension using a visual world eye-tracking paradigm. L2 learners who were exposed to thematic lists (TL) (e.g., *travel*) showed, initially, less improvement than learners who were exposed to visual scenes (VS) with vocabulary related by co-occurrence (e.g., *avión-bikini*, 'airplane-bikini'). This improvement was driven by higher anticipatory looks to the target upon hearing a biasing verb (e.g., *subir* 'to board') when there was no other plausible target (e.g., *subir* + *avión* 'to board

+ airplane' vs. *subir* + *hueso* 'to board + bone'). In cases where the target and distractor were semantically related, and equally plausible to finish the sentence (e.g., *avión-vuelo* 'airplane-flight'), participants showed semantic interference from the pretest (before training) and very little improvement throughout the training.

The improvements in the VS group compared to the TL group suggest that the VS training was closer to the naturalistic exposure received by native speakers and may help create stronger connections in the lexicon. The mechanisms behind this improvement cannot be fully explained in this study, and further research is needed to shed light on the psychological underpinnings behind L2 vocabulary learning and semantic network development. This paper is a step in the right direction towards creating better instructional materials for L2 classroom instruction that would help learners improve the efficiency of their lexicon for communication.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/languages9020043/s1>, Supplementary S1: Description of Category Fluency Task; Supplementary S2: Description of Digit Span Task; Supplementary S3: Process for calculating semantic relatedness scores within each visual world display; Supplementary S4: Procedure for generating stimuli through a word association task. Supplementary S5: Output of additional linear mixed-effects models for visual world data.

Funding: This research was funded by the 2021 Duolingo Dissertation Grant in Language Learning with Technology (Duolingo Inc.). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of The Pennsylvania State University (protocol code STUDY00016649 approved on 21 December 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ongoing post hoc analyses.

Acknowledgments: The author would like to thank Matthew T. Carlson and Paola E. Dussias for their support and guidance. She also acknowledges the valuable comments of the anonymous reviewers that substantially improved this manuscript. The pilot study for this work was supported by the Iguacu Excellence Award Honoring Carmen and Laurentino Gomes (Department of Spanish, Italian, and Portuguese at The Pennsylvania State University). The preparation of this manuscript was supported by the Humanities Scholarship Enhancement Fund (College of Liberal Arts and Sciences at the University of Florida).

Conflicts of Interest: The author declares no conflicts of interest.

References

- Altmann, Gerry T. M., and Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73: 247–64. [CrossRef]
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78. [CrossRef]
- Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015. Parsimonious mixed models. *arXiv* arXiv:1506.04967.
- Baus, Cristina, Albert Costa, and Manuel Carreiras. 2013. On the effects of second language immersion on first language production. *Acta Psychologica* 142: 402–9. [CrossRef]
- Beatty-Martínez, Annie L., Paola E. Dussias, Rosa E. Guzzardo Tamargo, Christina A. Navarro-Torres, María Teresa Bajo, and Judith F. Kroll. 2020. Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46: 1022–47. [CrossRef]
- Boersma, Paul, and David Weenink. 2017. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.36. Available online: <http://www.praat.org> (accessed on 23 January 2024).
- Bolger, Patrick, and Gabriela Zapata. 2011. Semantic categories and context in L2 vocabulary learning. *Language Learning* 61: 614–46. [CrossRef]
- Bordag, Denisa, Kira Gor, and Andreas Opitz. 2022. Ontogenesis Model of the L2 lexical representation. *Bilingualism: Language and Cognition* 25: 185–201. [CrossRef]

- Bousfield, W. A. 1953. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology* 49: 229–40. [\[CrossRef\]](#)
- Bybee, Joan. 1998. The emergent lexicon. *Chicago Linguistic Society* 34: 421–35.
- Cancho, Ramón Ferrer i, and Ricard V. Solé. 2001. The small world of human language. *Proceedings of the Royal Society of London* 268: 2261–65. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cofer, Charles N. 1966. Some evidence for coding processes derived from clustering in free recall. *Journal of Verbal Learning and Verbal Behavior* 5: 188–92. [\[CrossRef\]](#)
- Cohen, Burton H. 1963. Recall of categorized word lists. *Journal of Experimental Psychology* 66: 227–34. [\[CrossRef\]](#)
- Collins, Allan M., and Elizabeth F. Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological Review* 82: 407–28. [\[CrossRef\]](#)
- Connell, Katrina, M. Gabriela Puscama, Joana Pinzon-Coimbra, Julia Rembalsky, Gloria Xu, Jorge R. Valdés Kroff, María Teresa Bajo Molina, and Paola E. Dussias. 2021. Phonologically cued lexical anticipation in L2 English: A visual world eye-tracking study. In *Proceedings of the 45th Annual Boston University Conference on Language Development*. Boston: Cascadia Press.
- Craik, Fergus I. M., and Endel Tulving. 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104: 268–94. [\[CrossRef\]](#)
- Craik, Fergus I. M., and Robert S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior* 11: 671–84. [\[CrossRef\]](#)
- Cubillos, Jorge H. 2015. *Charlemos: Conversaciones Prácticas*. Boston: Pearson.
- Davies, Mark. 2016–. Corpus del Español: Two Billion Words, 21 Countries. Available online: <http://www.corpusdelespanol.org/web-dial> (accessed on 4 August 2021).
- De Deyne, Simon, Daniel J. Navarro, Amy Perfors, Marc Brysbaert, and Gert Storms. 2019. The “small world of words” English word association norms for over 12,000 cue words. *Behavior Research Methods* 51: 987–1006. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Deyne, Simon, Daniel J. Navarro, and Gert Storms. 2013. Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods* 45: 480–98. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Deyne, Simon, and Gert Storms. 2008. Word associations: Network and semantic properties. *Behavior Research Methods* 40: 213–31. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Deyne, Simon, and Gert Storms. 2014. Word associations. In *The Oxford Handbook of the Word*. Edited by John R. Taylor. Oxford: Oxford University Press, pp. 465–80.
- DeLong, Katherine A., Thomas P. Urbach, and Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8: 1117–21. [\[CrossRef\]](#)
- Dijkgraaf, Aster, Robert J. Hartsuiker, and Wouter Duyck. 2017. Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition* 20: 917–30. [\[CrossRef\]](#)
- Dijkgraaf, Aster, Robert J. Hartsuiker, and Wouter Duyck. 2019. Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience* 34: 881–900. [\[CrossRef\]](#)
- Dubossarsky, Haim, Simon De Deyne, and Thomas T. Hills. 2017. Quantifying the structure of free association networks across the life span. *Developmental Psychology* 53: 1560–70. [\[CrossRef\]](#)
- Duchon, Andrew, Manuel Perea, Nuria Sebastián-Gallés, Antonia Martí, and Manuel Carreiras. 2013. EsPal: One-stop shopping for Spanish word properties. *Behavioral Research Methods* 45: 1246–58. [\[CrossRef\]](#)
- Dussias, Paola E., Jorge R. Valdés Kroff, Rosa E. Guzzardo Tamargo, and Chip Gerfen. 2013. When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition* 35: 353–87. [\[CrossRef\]](#)
- Finkbeiner, Matthew, and Janet Nicol. 2003. Semantic category effects in second language word learning. *Applied Psycholinguistics* 24: 369–83. [\[CrossRef\]](#)
- Gilmore, Alex. 2007. Authentic materials and authenticity in foreign language learning. *Language Teaching* 40: 97–118. [\[CrossRef\]](#)
- González-Aguilar, María, and Marta Rosso-O’Laughlin. 2005. *Atando Cabos: Curso Intermedio de Español*, 2nd ed. Boston: Pearson.
- Grüter, Theres, Casey Lew-Williams, and Anne Fernald. 2012. Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research* 28: 191–215. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hallett, Peter E. 1986. Eye movements and human visual perception. *Handbook of Perception and Human Performance* 1: 10–11.
- Hart, Betty, and Todd R. Risley. 1992. American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology* 28: 1096–105. [\[CrossRef\]](#)
- Hills, Thomas, Mounir Maouene, Josita Maouene, Adam Sheya, and Linda Smith. 2009. Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Journal of the Association for Psychological Science* 20: 729–39. [\[CrossRef\]](#)
- Hopp, Holger. 2013. The development of L2 morphology. *Second Language Research* 29: 3–6. [\[CrossRef\]](#)
- Hsiao, Indy Y. T., Yu-Ju Lan, Chia-Ling Kao, and Ping Li. 2017. Visualization analytics for second language vocabulary learning in virtual worlds. *Educational Technology & Society* 20: 161–75.
- Huettig, Falk. 2015. Four central questions about prediction in language processing. *Brain Research* 1626: 118–35. [\[CrossRef\]](#) [\[PubMed\]](#)
- Huettig, Falk, and Gerry T. M. Altmann. 2005. Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition* 96: B23–B32. [\[CrossRef\]](#) [\[PubMed\]](#)

- Huetting, Falk, and Nivedita Mani. 2016. Is prediction necessary to understand language? Probably not. *Language, Cognition, and Neuroscience* 31: 19–31. [CrossRef]
- Huttenlocher, Janellen, Heidi Waterfall, Marina Vasilyeva, Jack Vevea, and Larry V. Hedges. 2010. Sources of variability in children's language growth. *Cognitive Psychology* 61: 343–65. [CrossRef] [PubMed]
- Hwang, Jin Kyoung, Jeannette Mancilla-Martinez, Janna Brown McClain, Min Hyun Oh, and Israel Flores. 2020. Spanish-speaking English learners' English language and literacy skills: The predictive role of conceptually-scored vocabulary. *Applied Psycholinguistics* 41: 1–24. [CrossRef]
- Ito, Aine, Martin J. Pickering, and Martin Corley. 2018. Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language* 98: 1–11. [CrossRef]
- Jackson, Alice F., and Donald J. Bolger. 2014. Using a high-dimensional graph of semantic space to model relationships among words. *Frontiers in Psychology* 5: 385. [CrossRef]
- Jiménez Jiménez, Antonio F. 2010. A comparative study on second language vocabulary development: Study abroad vs. classroom settings. *Frontiers: The Interdisciplinary Journal of Study Abroad* 19: 105–24. [CrossRef]
- Kaan, Edith. 2014. Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism* 4: 257–82. [CrossRef]
- Kaan, Edith, Andrea Dallas, and Frank Wijnen. 2010. Syntactic predictions in second-language sentence processing. In *Structure Preserved: Studies in Syntax for Jan Koster*. Edited by C. Jan-Wouter Zwart and Mark de Vries. Amsterdam: John Benjamins, pp. 207–13.
- Kaan, Edith, and Theres Grüter. 2021. Prediction in second language processing and learning: Advances and directions. In *Prediction in Second Language Processing and Learning*. Edited by Edith Kaan and Theres Grüter. Amsterdam: John Benjamins, pp. 1–24.
- Koizumi, Rie, and Yo In'nami. 2013. Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research* 4: 900–13. [CrossRef]
- Kuperberg, Gina R., and T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition, and Neuroscience* 31: 32–59. [CrossRef] [PubMed]
- Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2016. Tests in Linear Mixed Effects Models. Version 2.0.32. Available online: <https://cran.r-project.org/web/packages/lmerTest/index.html> (accessed on 15 April 2017).
- Lew-Williams, Casey, and Anne Fernald. 2007. Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science* 18: 193–98. [CrossRef] [PubMed]
- Lew-Williams, Casey, and Anne Fernald. 2010. Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language* 63: 447–64. [CrossRef] [PubMed]
- Linck, Jared A., Judith F. Kroll, and Gretchen Sunderman. 2009. Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science* 20: 1507–15. [CrossRef] [PubMed]
- Llanes, Àngels, and Carmen Muñoz. 2009. A short stay abroad: Does it make a difference? *System* 37: 353–65. [CrossRef]
- López-Beltrán Forcada, Priscila. 2021. Heritage Speakers' Online Processing of the Spanish Subjunctive: A Comprehensive Usage-Based Study. Doctoral dissertation, Electronic Thesis and Dissertations for Graduate School, The Pennsylvania State University, State College, PA, USA.
- Lumen Learning. 2018–2019. Introductory Spanish II. Available online: <https://courses.lumenlearning.com/wm-spanish2> (accessed on 15 April 2022).
- Luo, Lin, Gigi Luk, and Ellen Bialystok. 2010. Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition* 114: 29–41. [CrossRef]
- Malt, Barbara C. 2010. Naming artifacts: Patterns and processes. In *The Psychology of Learning and Motivation: Advances in Research and Theory*. Edited by Brian H. Ross. Amsterdam: Elsevier, pp. 1–38.
- Marian, Viorica, and Margarita Kaushanskaya. 2007. Cross-linguistic transfer and borrowing in bilinguals. *Applied Psycholinguistics* 28: 369–90. [CrossRef]
- Meara, Paul. 1996. The dimensions of lexical competence. In *Performance and Competence in Second Language Acquisition*. Edited by Gillian Brown, Kirsten Malmkjaer and John Williams. Cambridge: Cambridge University Press, pp. 33–55.
- Milton, James, and Paul Meara. 1995. How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics* 107: 17–34. [CrossRef]
- Mir, Montserrat, and Ángela Bailey de las Heras. 2015. *¡Qué me Dices! A Task-Based Approach to Spanish Conversation*. Boston: Pearson.
- Nagy, William E., and Patricia A. Herman. 1987. Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In *The Nature of Vocabulary Acquisition*. Edited by Margaret G. McKeown and Mary E. Curtis. Mahwah: Lawrence Erlbaum Associates, Inc., pp. 19–35.
- Olivella de Castells, Matilde, Elizabeth E. Guzmán, P. Paloma Lapuerta, and Judith E. Liskin-Gasparro. 2013. *Mosaicos: Course Materials for Spanish 2*, 4th Custom ed. Boston: Pearson.
- Olivella de Castells, Matilde, Elizabeth E. Guzmán, P. Paloma Lapuerta, and Judith E. Liskin-Gasparro. 2015. *Mosaicos: Course Materials for Spanish 3*, 5th Custom ed. Boston: Pearson.
- Pizziconi, Barbara. 2017. Japanese vocabulary development in and beyond study abroad: The timing of the year abroad in a language degree curriculum. *The Language Learning Journal* 45: 133–52. [CrossRef]
- Potowski, Kim, S. Silvia Sobral, and Laila M. Dawson. 2012. *Dicho y Hecho: Beginning Spanish*, 9th ed. Hoboken: John Wiley & Sons.

- Prolific. 2014. Available online: <https://www.prolific.co> (accessed on 18 November 2021).
- Qian, David D. 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning* 52: 513–36. [CrossRef]
- Qualtrics. 2005. Provo, UT. Available online: <https://www.qualtrics.com> (accessed on 18 November 2021).
- R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <http://www.R-project.org> (accessed on 15 November 2023).
- Recchia, Gabriel, and Michael N. Jones. 2009. More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods* 41: 647–56. [CrossRef]
- SR Research Data Viewer 3.1.246 [Computer Software]. 2004–2015a. Mississauga: SR Research Ltd.
- SR Research Experiment Builder 2.2.38 [Computer Software]. 2004–2015b. Mississauga: SR Research Ltd.
- Stæhr, Lara S. 2009. Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition* 31: 577–607. [CrossRef]
- Story Board That. 2022. Available online: <https://www.storyboardthat.com> (accessed on 18 November 2021).
- Tamariz, Mónica. 2000. *Oxford Spanish Cartoon-Strip Vocabulary Builder*. Oxford: Oxford University Press.
- Terrell, Tracy, Magdalena Andrade, Jeanne Egasse, and Elías M. Muñoz. 2001. *Dos Mundos*, 5th ed. New York: McGraw-Hill.
- Voeten, Cesko C. 2020. buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression. Available online: <https://cran.r-project.org/web/packages/buildmer/index.html> (accessed on 1 April 2022).
- Zaytseva, Victoria, Carmen Pérez-Vidal, and Imma Miralpeix. 2018. Vocabulary acquisition during study abroad: A comprehensive review of the research. In *The Routledge Handbook of Study Abroad Research and Practice*. Edited by Cristina Sanz and Alfonso Morales-Front. New York: Routledge, pp. 210–24.
- Zhao, Xiaowei, and Ping Li. 2010. Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism* 13: 505–24. [CrossRef]
- Zoom Video Communications. 2022. Available online: <https://zoom.us> (accessed on 23 January 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.